RESEARCH ON COMPRESSION ENHANCEMENT USING LEARNING BASED APPROACH AND ADDITIONAL INFORMATION

著者	Trinh Man Hoang
出版者	法政大学大学院理工学・工学研究科
journal or publication title	法政大学大学院紀要. 理工学・工学研究科編
volume	62
page range	1-4
year	2021-03-24
URL	http://doi.org/10.15002/00023950

RESEARCH ON COMPRESSION ENHANCEMENT USING LEARNING BASED APPROACH AND ADDITIONAL INFORMATION

18R8104 Trinh Man Hoang

Abstract— **Recently, learning-based methods have been well known for their efficiency in the multimedia enhancement task. In this work, we studied the ability of learning-based methods in the compression enhancement. In detail, a special design of Recursive Residual Neural network (RRN) is applied to enhance the conventional HEVC video codec – B-DRRN, a layered image compression uses our SMapNet for the semantic segment enhancement - EDMS and a hybrid video codec uses self-enhancement and multi-frames enhancement. We also conduct a new large-scale dataset with 209,152 training samples. Experimental results show that the proposed B-DRRN and the hybrid codec can reduce up to 6.16% and 33.73% BD-rate compares to HEVC and x265, respectively. Whereas, the proposed EDMS can get 5% bitrate, and 24% encoding time saving compare to the state-of-the-art semantic-based image codec.**

I. INTRODUCTION

The typical lossy compression standards and codecs are mostly processed based on block-wise transformation and quantization. However, it also leads to extreme blurring and block-type artifacts. There are three main approaches to apply learning methods for the compression enhancement task. The first direction is to replace those components of the conventional codec with deep tools. The second direction is to perform an end-to-end learning manner for the whole codec. Third, both conventional and end-to-end methods can be hybrid used to offset each other.

In recent years, the additional information has been proved as a huge impulse for the compression enhancement task. Even so, using additional information required a bigger file size on the bitstream and more complex models as a tradeoff for the media quality enhancement. To overcome this tradeoff, we proposed several learning-based designs towards the mentioned approaches. In summary, our contributions are mainly 4 folds:

- We develop a novel layered image compression framework – EDMS.
- We propose a hybrid video compression framework that employs the cross-enhancement.
- We design a novel network architecture using the Recursive Residual structure - B-DRRN to leverage the block information of the coding unit (CU).

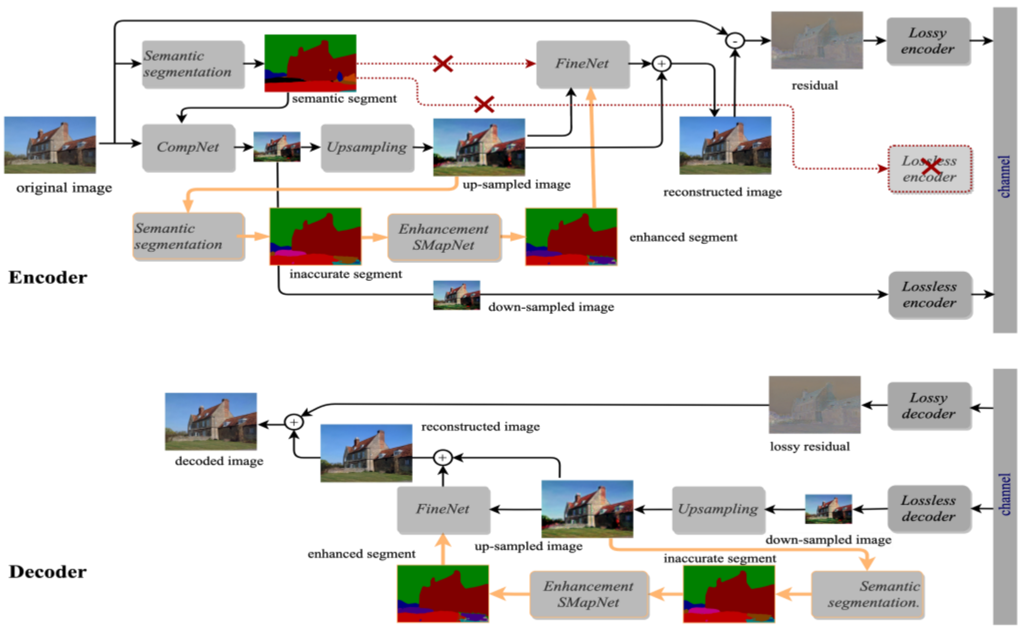
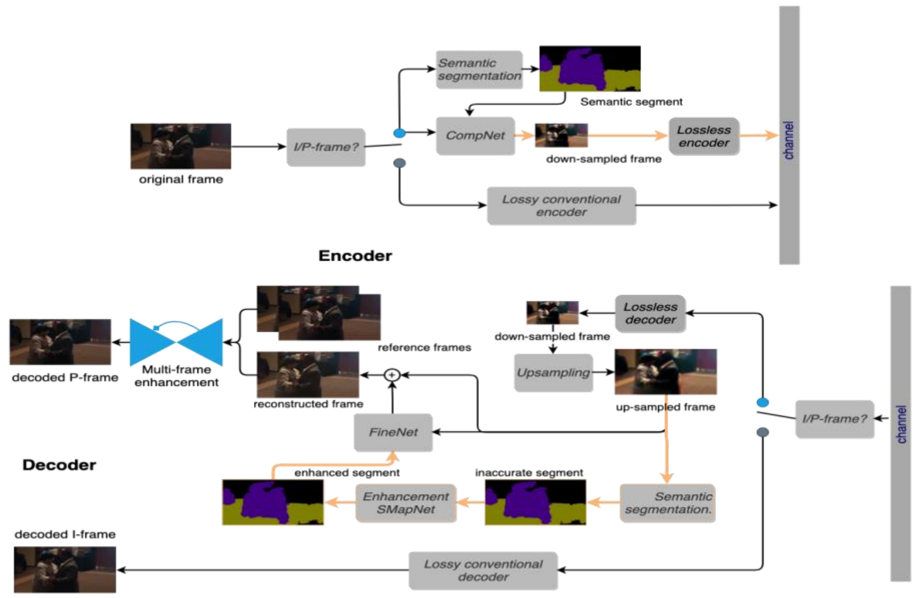


Figure 1. EDMS overall framework.



- Moreover, we conducted a new large-scale dataset with 209,152 training samples for training such deep CNN as B-DRRN.

II. METHODS

A. Image compression with Encoder-Decoder Matched Semantic segmentation – EDMS

A.1. Encoder-decoder matched semantic segmentation

Akbari *et al* [1] introduced a CompNet at the encoder to down-sample the original image to a compact version. Recently, a GAN-based FineNet [2] was also demonstrated that it can generate a synthesis image from the up-sampled version and the semantic segment. We extract the semantic segment from this up-sampled version for both encoder-decoder. However, it is not as good as the one extracted from the original image. Hence, we applied the Recursive Residual architecture [3] as a mapping operator - SMapNet. The residual now is calculated based on the synthesis image which conducted from the enhanced segment and the up-sampled image.

A.2. EDMS overall framework

Figure 1 shows our EDMS overall framework. On the encoder side, we first perform the down-sampling and up-sampling processes with the original semantic segment. Next step, we extract the segment from the up-sampled version and use the SMapNet for semantic segmentation enhancement. The final residual will be calculated based on the output of FineNet forward with SNetMap segment as its input. On the decoded side, we received only the down-sampled image and lossy residual from the channel. The semantic segment used to reconstruct the decoded image is conducted from the up-sampled image and enhanced by our SMapNet. Next, FineNet uses this enhanced segment and the up-sampled image as its input to perform the reconstruction. The reconstructed image is then sum up with the residual to output the final decoded image.

B. Hybrid video compression framework

B.1. P-frame self-enhancement

In our hybrid enhancement framework, with each video sequence, two separate bitstreams will be sent, one for the conventional I-frames and another one for the learning-based P-frames. All frames between two consecutive I-frames, including themselves, are formed as a Group-Of-Frame (GOF). For the P-frame sending information, we see that among two sending information of the EDMS, the down-sampled version is perfect fit for ours model.

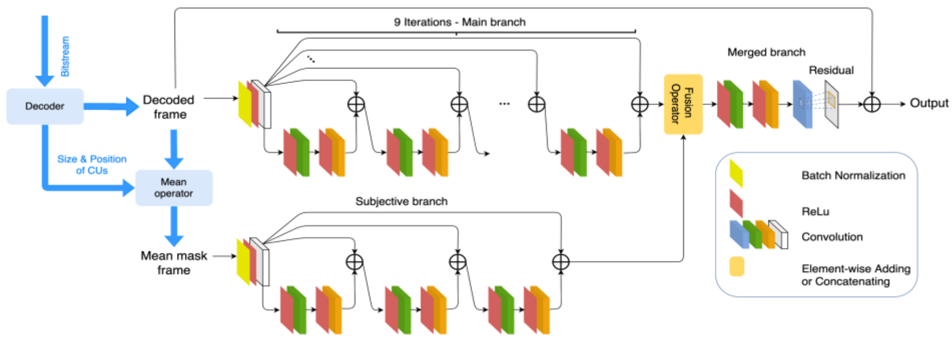


Figure 3. B-DRRN overall architecture.

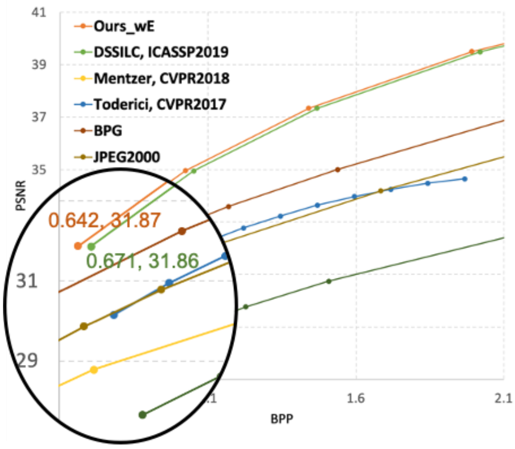


Figure 4. The comparison of compression techniques using PSNR.

B.2. Multi-frames enhancement

The input of this module is a triplet of I-P self-enhanced-I frames, we calculated the optical flow by input the self-enhanced P-frame and its references. Then these flows will be backward warped, in combination with the visible map and U-Net architecture, our model can enhance the P-frame by transfer the texture from the I-frames.

B.3. Hybrid video compression overall framework

Figure 2 shows our overall hybrid video compression framework. On the encoder side, we first perform the down-sampling with the original semantic segment. On the decoded side, we received only the down-sampled frame. The semantic segment used to reconstruct the decoded image is conducted from the up-sampled image and enhanced by our SMapNet. Next, FineNet uses this enhanced segment and the up-sampled image to perform the reconstruction. The reconstructed image then acts as a P-frame input for the multi-frames' enhancement model.

C. Block information constrained Deep Recursive Residual Network – B-DRRN

C.1. Extra branch with block information

The artifact of block-based video coding standards comes from the block dividing process. Recently, the Mean mask conducted from this block information was proven that could get a better result than some other approaches by Xiaoyi He[4]. In this work, we use a similar representation and apply the Recursive Residual block from DRRN [5] for this additional branch. The output of our two branches then will be fusion. After the fusion, we design a merged branch contains two residual layers with the role of the noise-canceling and feature reordering unit. Furthermore, to keep intact parameters, we implement the technique of sharing weight for all the parameters among all branches of our new model.

C.2. B-DRRN overall architecture

Figure 3 shows our overall architecture. To process a stable training, a batch normalization layer is first applied for each input and its correlative block mean mask. The first convolution layer contains 64 filters with a size of 3x3. And the Recurrent Residual Unit also has two similar convolution layers. At the main branch, the Recurrent Residual unit repeats nine times and to keep the balance between two branches, while the extra branch will have three iterations. The outputs of these two branches are combined by adding operator or concatenating operator. For concatenating fusion, the output dimension of the fusion layer will be double, so an additional convolution layer goes after to reduce the dimension of the feature from 128 to 64. Then, our merge branch is applied. After that, a 3x3 filter is used for residual reconstruction. Finally, the residual image is added directly to the input. In our new model, the Rectifier Linear Unit - ReLu is applied before doing any convolutional calculation.

III. RESULTS

A. EDMS

We compare our method with some learning-based and several traditional methods. The RD-curve are shown in Figure 4. From Figure 4, we can observe that our method still achieves the upper-bound in the consideration of PSNR. In particular, our method gains 35.31% BD-rate reduction over BPG.

Table 1. The RD-rate reduction over x265 codec with AI-configuration on HM common test sequences.

Sequences	**BD-rate reduction (%)**		
	GOF = 3	**GOF = 5**	**GOF = 9**
Traffic	-21.29	-32.49	-37.26
BQTerrace	-12.25	-16.79	-22.45
ParkSence	-17.55	-33.31	-26.14
BasketballDrill	-12.34	-29.91	-7.78
BQMall	-17.67	-29.56	-14.41
PartyScene	-28.60	-49.44	-60.31
BQSquare	-26.21	-50.25	-65.25
BlowingBubbles	-29.41	-52.14	-64.15
BasketballPass	-11.55	-21.27	-14.7
RaceHorces D	-16.14	-21.84	9.53
Johnny	-7.31	-21.64	-29.22
FourPeople	-18.68	-43.06	-56.26
Kristen&Sara	-14.08	-36.76	-45.04
Average	**-17.93**	**-33.73**	**-33.34**

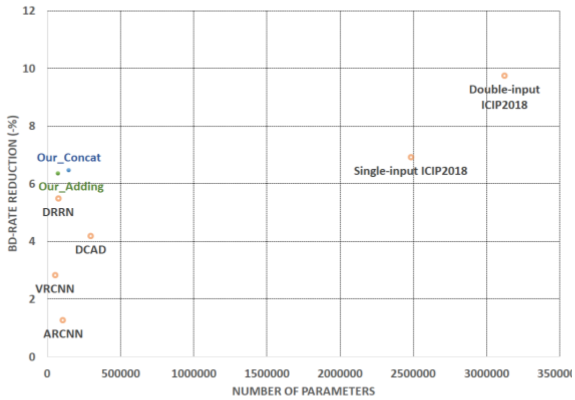


Figure 5. Our proposal Our_Adding could reduce more BD-rate without adding more parameters.

B. *Hybrid video enhancement*

Video compression performance. Table 1 shows the BD-rate reduction over the x265 all-intra configuration. We can first observe that our approach is better than x265 for most of the cases. In particular, we can get up to 65% BD-rate reduction compare to x265.

GOF selection. To evaluate the effect of our Group-Of-Frame selection, we conducted some experimented on three GOF = 3, 5, and 9. As shown in Table 1, we can clearly see that the performance mainly increased when the GOF increase.

C. *B-DRRN*

Figure 5 shows the BD-rate reduction of some deep tools, the baseline model DRRN[5] and Our models. In Figure 5, our approaches (Our_Concat and Our_Adding) achieve better performance than DRRN when comparing to HM-20.0 software, our approaches can get 6.48% and 6.37% BD-rate reduction respectively. Figure 5 also shows the relationship between the number of parameters and the BD-rate reduction (-%) of all competitive models. By applying B-DRRN, there is a big gap in BD-rate reduction between Our_Adding and the baseline DRRN but it remains the same on the number of parameters. Compare to the Double-input CNN, although it gets a higher BD-rate reduction than Ours, it needs 38x more parameters compare to ours.

IV. CONCLUSION

This work presents several methods that using additional information to enhance the compression quality. In particular, we proposed an Encoder-Decoder Match Semantic segmentation – EDMS framework for layered image compression, a Hybrid video compression framework with P-frame enhancement models, and a Block-information constrained Deep Recursive Residual Network – B-DRRN for the compression post-processing.

Experimental results showed that the proposed approaches could outperform all convention multimedia codecs. Since there still is a lot of space for framework improving and much information can be synchronously extracted from both encoder and decoder, our approaches have the potential to be extended for other future work.

ACKNOWLEDGMENT

I'm extremely grateful to my supervisor, Professor Jinjia Zhou, who has provided me with encouragement and patience throughout the duration of this project. I would also like to extend my deepest gratitude to Professor Kazuo Yana, who has brought me the opportunity to experience and study at Hosei University. Many thanks all Intelligent Media Processing Lab's members for their assists and sharing knowledge. My thanks and appreciation also go to people who have willingly helped me out with their ability.

REFERENCES

[1] M. Akbari, J. Liang, and J. Han, "DSSLIC: Deep Semantic Segmentation-based Layered Image Compression," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 2042–2046, doi: 10.1109/ICASSP.2019.8683541.

[2] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, Jun. 2018, pp. 8798–8807, doi: 10.1109/CVPR.2018.00917.

[3] Y. Tai, J. Yang, and X. Liu, "Image Super-Resolution via Deep Recursive Residual Network," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 2790–2798, doi: 10.1109/CVPR.2017.298.

[4] X. He, Q. Hu, X. Zhang, C. Zhang, W. Lin, and X. Han, "Enhancing HEVC Compressed Videos with a Partition-Masked Convolutional Neural Network," in 2018 25th IEEE International Conference on Image Processing (ICIP), Oct. 2018, pp. 216–220, doi: 10.1109/ICIP.2018.8451086.

[5] Y. Tai, J. Yang, and X. Liu, "Image Super-Resolution via Deep Recursive Residual Network," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 2790–2798, doi: 10.1109/CVPR.2017.298.